新农合药品报销目录跨省整合信息系统设计与实施*

李亚子　郑见立　周奕洋　李国垒
(中国医学科学院医学信息研究所　北京　100020)

摘要:【目的】分析新农合药品报销目录存在的问题,提出整合技术路线,开发信息系统,整合多来源词表,生成国家新农合药品一体化目录。【方法】借鉴 UMLS 整合技术路线,采用映射的方法对多来源药品字典集成存在的问题进行解决。【结果】制定了相应的新农合药品编码目录数据结构和映射算法,完成了国家新农合一体化目录构建系统的设计与开发。【局限】药品之间相互药理作用关系还需进一步梳理完善。【结论】经过制定数据结构和映射算法对新农合目录编码进行映射,解决了构建国家新农合药品目录编码中多来源字典集成问题。

关键词: 新型农村合作医疗　多来源字典　药品编码　集成映射
分类号: R19　G350

1 引　言

随着新农合政策的执行,实现数据交换和数据共享已成为大势所趋,数据的共享有助于推动新农合政策的贯彻落实,实现新农合患者就医便利、报销便捷[1]。其中新农合药品报销目录是进行就医结算报销的依据,但是目前各省新农合信息系统药品目录不一致,难以进行相关数据分析以及跨省就医即时结算。因此有必要建立统一性、延展性与适应性均较好的国家一体化药品目录。

在对多个异构数据源进行集成时,若数据集成系统中的本体是单独建立而不是参照同一个共享本体构建,会造成局部本体间存在语义不一致的情况,为了系统内查询处理的正常运行,必须建立本体映射[2]。目前国内外就异构数据源集成技术已展开较多的研究。国外方面,Husain 等开发 SOCR (Statistics Online Computational Resource)系统来处理不同来源的数据集成问题[3]; Verbeke 等提出基于语义网络的数据集成方法实现对不同来源的肿瘤数据的集成[4]; Scheurwegs 等对每一个数据源分别建立模型,然后利用 Meta-Learner 分别对模型进行评价,最终实现异构数据源的集成[5]。在国内方面,才苗提出基于神经网络的异构数据库语义集成技术[6]; 文语欣则采用语义技术解决临床信息异构问题,提出基于物理层、语义层和应用层的三层架构模型的临床信息交换框架,解决医疗信息系统之间的信息共享[7]; 郭鑫提出基于本体的异构数据集成技术,以解决异构数据源之间存在的系统异构、语法异构和语义异构问题[8]。

UMLS(Unified Medical Language System)[9]是使用最为广泛的计算机化的情报检索语言集成系统,作为语言规范化、翻译的工具,可实现跨数据库交互的词汇转换。UMLS 通过对术语的标化等处理[10],实现了对 100 余词典 100 万以上的概念的集成[11],应用于健康信息互操作、医学术语、药品名称的联接、保险账单的编码等。本文主要针对新农合药品报销目录编码进行集成,由于各省的目录编码都是在《国家基本医疗保险、工伤保险和生育保险药品目录(2009 版)》、《国家基本药物目录(2012 版)》目录的

通讯作者: 李亚子, ORCID: 0000-0002-6669-2041, E-mail: li.yazi@imicams.ac.cn。
*本文系国家社会科学基金项目"医学知识组织体系映射模式及其在医保中的应用服务研究"(项目编号:13CTQ032)的研究成果之一。

基础上根据本地情况进行适当改造[12]。因此本文借鉴 UMLS 的技术整合路线，参考 UMLS 词表结构进行数据结构设计，完成了《国家基本药物目录》、《社会基本医疗保险、工伤保险和生育保险药品目录》以及安徽、河南等 10 个省的新农合药品报销目录的整合。

2　研究对象与方法

2.1　数据来源

来源词表包括：《国家基本药物目录(2012 版)》(简称“国家基药目录”)(药品数，520)、《国家基本医疗保险、工伤保险和生育保险药品目录(2009 版)》(简称“社保目录”)(药品数，3 277)，以及安徽(药品数，1 885)、福建(药品数, 2 341)、河南(药品数, 2 714)、江苏(药品数, 1 198)、四川(药品数, 1 330)、河北(药品数, 1 686)、吉林(药品数, 1 806)、黑龙江(药品数, 2 172)、湖南(药品数, 1 781)、内蒙古(药品数, 2 082)等 10 省新农合药品报销目录。

2.2　研究方法

(1) 数据结构设计

本文借鉴 UMLS 词表整合系统设计了国家新农合一体化目录构建系统数据结构，如图 1 所示:

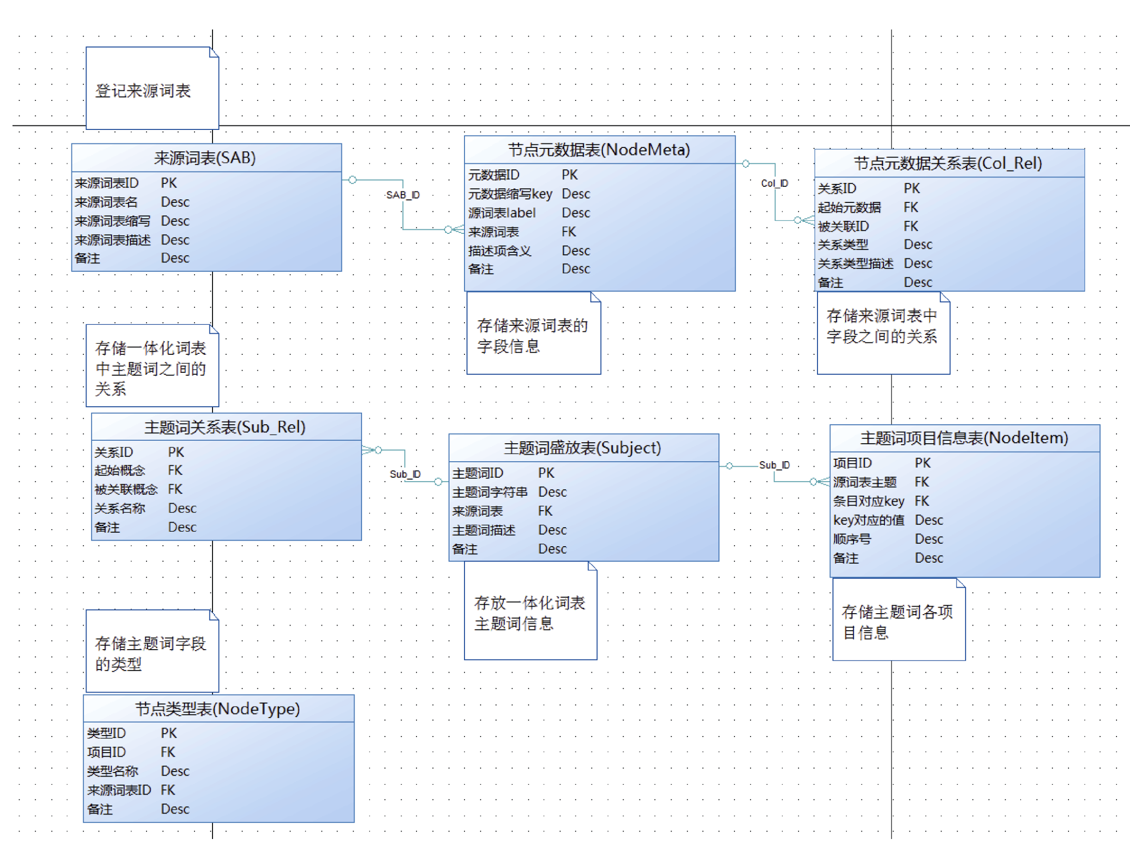


图 1　国家新农合一体化目录构建系统的核心数据结构

图1中来源词表(SAB)用于登记来源词表，将国家基药目录、社保目录以及安徽等10个省目录按顺序存放到来源词表中；节点元数据表(NodeMeta)用于存放来源字表各字段信息，将国家基药目录、社保目录以及各省级药品目录中的字段信息存放到节点元数据表中；节点元数据关系表(Col_Rel)用于存储节点元数据表中各个字段之间的关系；主题词盛放表(Subject)用于存放一体化词表中的主题词相关信息；主题词项目信息表(NodeItem)用于存放主题词各项目详细信息；主题词关系表(Sub_Rel)用于存放一体化词表中主题词之间的关系；节点类型表(NodeType)用于存储主题词字段的类型，解决词系词簇问题。

(2) 映射算法设计

构建映射算法规则，实现各省级目录向国家级目

录集成映射，算法映射规则如下：

①提示与该节点的父节点相同或相似的子节点；

②提示与该节点某些字段相同的词条，如名称相同的，编码相同的，对于编码相同的，又可提示该编码的兄弟节点；

③提示与该节点相似度最大的前 20 个词条；

相似度计算 S=(字段 i 的相同部分^2/字段 i 的总长度) (A1 与 A2 相同部分^2/A1 字段与 A2 字段的长度之和)；

④在目录映射中，出现语义相同，名称不同的，主要原因是语序不一致，因此，忽略语序顺序进行词条比对相似度计算；

⑤提供包含关系的词条发现，如词条A，可检索完全包含词条A的词条。

(3) 映射流程

国家新农合一体化目录构建系统旨在基于各省新农合目录的基础上，构建一套全面、细粒度的国家级目录，从而建立各省与国家新农合一体化目录的映射。字典目录集成系统的构建流程如图 2 所示：

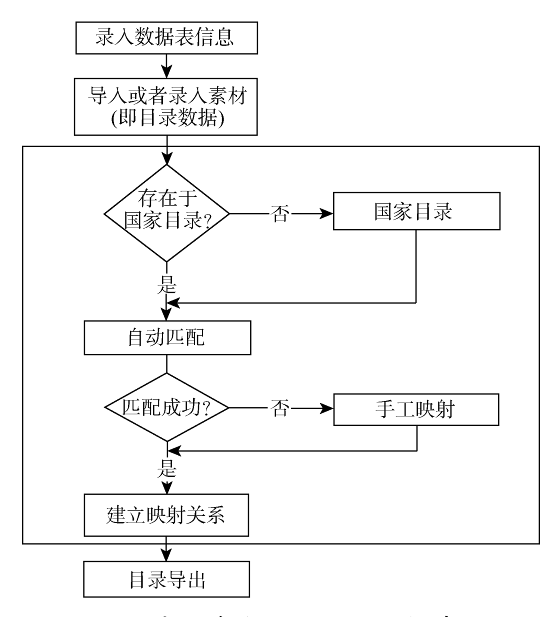


图 2　国家新农合一体化目录构建流程

首先将国家基药目录作为国家一体化目录的初始值，然后按顺序分别将社保目录，安徽、河南等各省级新农合目录转化为指定格式，根据制定的映射算法，完成药品名称的集成映射。

3　研究过程

3.1　整合实例

以 2012 版国家基药目录和 2015 年河北省药品目录中的“青霉素”为例阐述系统的映射过程。表 1 和表 2 分别为目录示例。

表 1　2012 版国家基药目录示例

序号	编码	品名	分类	标注剂型
1	GJJY-HS-1	青霉素	青霉素类	注射用无菌粉末

表 2　2015 年河北省药品目录示例

序号	编码	品名	拼音码	分类	标注剂型
1	10101006	青霉素	qmsjyzsywjfm	青霉素类	注射用无菌粉末

(1) 按顺序分别将国家基药目录、2015 年河北省药品目录登记到来源词表(SAB)中，存储形式如表 3 所示：

表 3　来源词表(SAB)存储示例

SAB_ID	SAB_Name	SAB_ABR	SAB_DESC	B1
2012 版国家基药目录 ID	国家基药目录	gjjyml	国家基药目录	2012
2015 年河北省药品目录 ID	2015 河北省药品目录	hbypml	河北省药品目录	2015

(2) 将来源词表中的各字段信息存放到节点元数据表(NodeMeta)中，存储形式如表 4 所示，并将来源词表字段之间的关系存放到节点元数据关系表(Col_Rel)中，在示例中并未涉及相关字段，但在江苏省 2009 年药物目录中存在县(市、区)和乡(镇)两个字段，该两个字段之间的隶属关系应存放到节点元数据关系表中。

(3) 将国家基药目录存放到主题词盛放表中作为初始值，形成初始化的国家新农合一体化目录，然后将社保目录，安徽、河南等新农合目录向一体化目录映射。在省级目录与国家新农合一体化目录进行映射时，将该省级目录与国家新农合一体化目录药品名称进行逐条比对，比对的对象是药品名称和剂型。如果国家新农合一体化药品目录中已包含该药品，则建立映射关系。如果国家新农合一体化目录中未包含该药品，则增加该药品到国家新农合一体化目录中。“是否包含”的判断依据为，品名一致或品名不一致但有效成分一致，并且剂型一致，则判断为“包含”。同一种药品，不同剂型视作不同记录，暂不考虑规格的差异。主题词盛放表示例如表 5 所示。由于 2015 年河北省药品目录“青霉素”药品向一体化目录映射时，该药品已经存在，且品名和剂型均一致，故不存放到主题词盛放表中，而 2015 年河北省

药品目录中的“头孢替唑”在一体化目录中不存在，故将其并增到一体化目录中。

(4) 将主题词盛放表中主题词的项目信息存放到主题词项目信息表(NodeItem)，如表 6 所示。

表 4　节点元数据表(NodeMeta)存储示例

Col_ID	Col_Key	LABEL	SAB_ID	Col_DESC	B1
序号 ID	xh	序号	2012 版国家基药目录 ID	基药目录字段	2012
编码 ID	bm	编码	2012 版国家基药目录 ID	基药目录字段	2012
品名 ID	pm	品名	2012 版国家基药目录 ID	基药目录字段	2012
分类 ID	fl	分类	2012 版国家基药目录 ID	基药目录字段	2012
剂型 ID	jx	标注剂型	2012 版国家基药目录 ID	基药目录字段	2012
序号 ID	xh	序号	2015 年河北省药品目录 ID	河北省药品目录字段	2015
编码 ID	bm	编码	2015 年河北省药品目录 ID	河北省药品目录字段	2015
品名 ID	pm	品名	2015 年河北省药品目录 ID	河北省药品目录字段	2015
拼音码 ID	pim	拼音码	2015 年河北省药品目录 ID	河北省药品目录字段	2015
分类 ID	fl	分类	2015 年河北省药品目录 ID	河北省药品目录字段	2015
标注剂型 ID	jx	标注剂型	2015 年河北省药品目录 ID	河北省药品目录字段	2015

表 5　主题词盛放表(Subject)存储示例

Sub_ID	Sub_STR	Sub_ID	Sub_DESC	B1
青霉素 ID	青霉素	2012 版国家基药目录 ID	青霉素药品	2012 版基药目录
头孢替唑 ID	头孢替唑	2015 年河北省药品目录 ID	头孢替唑	2015 年河北药品目录

表 6　主题词项目信息表(NodeItem)存储示例

Item_ID	Sub_ID	Item_Col_ID	Item_Col_Value	Node_Rank	B1
项目 ID1	2012 版国家基药目录 ID	xh	1	1	序号
项目 ID2	2012 版国家基药目录 ID	bm	GJJY-HS-1	2	编码
项目 ID3	2012 版国家基药目录 ID	pm	青霉素	3	品名
项目 ID4	2012 版国家基药目录 ID	fl	青霉素类	4	分类
项目 ID5	2012 版国家基药目录 ID	jx	标注剂型	5	剂型
项目 ID6	2015 年河北省药品目录 ID	xh	1	1	序号
项目 ID7	2015 年河北省药品目录 ID	bm	10102013	2	编码
项目 ID8	2015 年河北省药品目录 ID	pm	头孢替唑	3	品名
项目 ID9	2015 年河北省药品目录 ID	pim	tbtzzsj	4	拼音码
项目 ID10	2015 年河北省药品目录 ID	fl	头孢菌素	5	分类
项目 ID11	2015 年河北省药品目录 ID	jx	注射剂	6	剂型

将主题词之间的关系(即相关药物之间的协同(增效)、相加(增加)、拮抗(减效)等)存放到主题词关系表中，并将主题词所属的类型存放到节点类型表中，分别如表 7 和表 8 所示：

表 7　主题词关系表(Sub_Rel)存储示例

Sub_Rel_ID	From_Sub_ID	To_Sub_ID	Sub_Rel_Type	B1
关系 ID	胺碘酮 ID	美托洛尔 ID	协同作用	增效

表 8　节点类型表(NodeType)存储示例

NodeType_ID	Item_ID	Type_STR	SAB_ID	B1
类型 ID1	项目 ID4	青霉素类	2012 版国家基药目录 ID	2012 版基药目录

3.2　整合结果

通过对国家基药目录、国家社保药品目录、安徽等 10 省新农合药品目录进行整合。不考虑剂型的情况下，

共计4 136种，其中西药2 112种，中药2 024种；考虑剂型，共计6 061种，其中西药2 946种，中药3 115种。

4 系统核心功能实现

通过设计基于UMLS的数据结构，制定相关的映射算法，并以整合药品目录为基础，开发了国家新农合一体化目录构建系统，完成国家新农合目录(疾病编码、药品编码、诊疗编码规范、耗材编码规范和机构编码)的初始化及与各省新农合字典目录的对照，建立了关联，并实现了国家新农合一体化目录的及时更新。系统主要通过国家新农合一体化目录初始化和映射管理实现集成目录的构建。在目录初始化界面，选择一个目录大类，然后选择一个省份以及省目录版本，此时会在省目录框中出现所选择的目录树，再选择国家新农合一体化目录版本，国家新农合一体化目录框中会出现国家新农合一体化目录树，将左侧省目录树中的树节点拖动到右侧国家新农合一体化目录树下面的某个节点，即完成了国家新农合一体化目录的初始化功能，如图3所示：

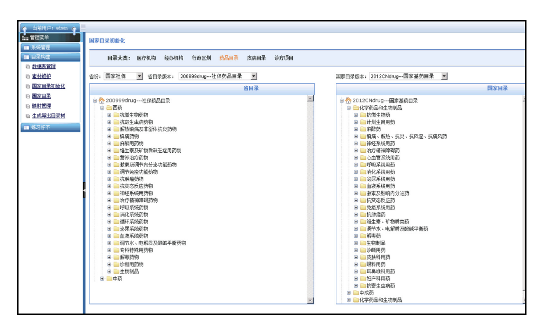


图3 国家新农合一体化目录初始化功能

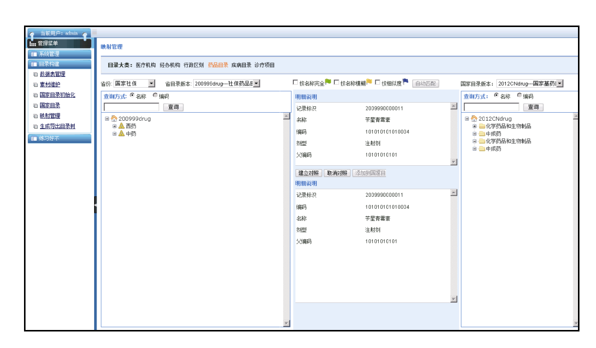


图4 映射管理功能

如图4所示，在映射管理功能模块中，选择一个目录大类，并选择一个省份以及省目录版本，和一个国家新农合一体化目录版本，在左侧显示的是省目录，右侧显示的是国家新农合一体化目录，然后点击建立对照，即可完成省级目录向国家新农合一体化目录的映射。

5 结 语

本文分析了构建新农合国家一体化目录中存在的问题，针对这些问题，提出了集成映射的解决方法，制定了相应的数据结构、集成映射规则与算法，并开发了国家新农合平台一体化目录构建系统，实现了各个省目录向国家新农合一体化目录的集成和映射。随着目录的成熟，以及配套政策的出台，可将集成目录用于新农合跨省就医的异地结算和即时结报中。

本研究还存在一定的局限，药品之间相互药理作用关系库还需进一步完善以及多种药品之间的药理作用关系还需进一步研究，系统功能完善后，将国家一体化目录作用于全国新农合跨省就医即时结报中，以指导各地结算政策的衔接。

参考文献：

[1] 王秋月. 新农合即报系统的研制与开发[J]. 中国数字医学, 2014(1): 109-111. (Wang Qiuyue. Research and Development of New Rural Cooperative Medical Instant Reimbursement System [J]. China Digital Medicine, 2014(1): 109-111.)

[2] 张红宇. 数据集成中本体映射的研究[D]. 长沙: 中南大学, 2005. (Zhang Hongyu. Research on Ontology Mapping in Data Integration [D]. Changsha: Central South University, 2005.)

[3] Husain S S, Kalinin A, Truong A, et al. SOCR Data Dashboard: An Integrated Big Data Archive Mashing Medicare, Labor, Census and Econometric Information [J]. Journal of Big Data, 2015. DOI: 10.1186/s40537-015-0018-z.

[4] Verbeke L P C. Van Den Eynden J, Fierro A C, et al. Pathway Relevance Ranking for Tumor Samples Through Network-Based Data Integration [J]. PLOS ONE, 2015, 10(7): e0133503.

[5] Scheurwegs E, Luyckx K, Luyten L, et al. Data Integration of Structured and Unstructured Sources for Assigning Clinical Codes to Patients Stays [J]. Journal of the American Medical Informatics Association, 2016, 23: e11-e19.

[6] 才苗. 基于神经网络的异构数据库语义集成的研究[D]. 大连: 大连理工大学, 2009. (Cai Miao. Research of Semantic Integration in Heterogeneous Database Based on Nerual

Network [D]. Dalian: Dalian University of Technology, 2009.)

[7] 文语欣. 基于语义的临床信息交换方法研究[D]. 杭州: 浙江大学, 2014. (Wen Yuxin. Study on Semantic-based Clinical Information Exchange Method [D]. Hangzhou: Zhejiang University, 2014.)

[8] 郭鑫. 基于本体的异构数据集成技术研究与实现[D]. 北京: 中国航天第二研究院, 2008. (Guo Xin. Research and Implementation of Ontology Based Heterogenous Data Integration Technology [D]. Beijing: The Second Academy of China Aerospace, 2008.)

[9] UMLS [EB/OL]. [2015-07-25]. http://www.nlm.nih.gov/research/umls/sourcereleasedocs/index.html.

[10] McCray A T, Srinivasan S, Browne A C. Lexical Methods for Managing Variation in Biomedical Terminologies[C]. In: Proceedings of the Annual Symposium on Computer Application in Medical Care. 1994: 235-239.

[11] UMLS 数据源介绍[EB/OL]. [2015-07-26]. http://www.nlm.nih.gov/research/umls/new_users/online_learning/OVR_001.html. (Introducition of UMLS Data Source [EB/OL]. [2015-07-26]. http://www.nlm.nih.gov/research/umls/new_users/online_learning/OVR_001.html.)

[12] 代涛, 李亚子, 郭珉江, 等. 新农合数据编码标准建设现状研究[J]. 中国卫生信息管理杂志, 2014, 11(1): 47-50, 56. (Dai Tao, Li Yazi, Guo Minjiang, et al. Study on Encoding Standard Application Status of the New Rural Cooperative Medical Data [J]. Chinese Journal of Health Informatics and Management, 2014, 11(1): 47-50, 56.)

作者贡献声明:

李亚子: 提出研究思路;
郑见立: 设计研究方法, 准备实验环境;
周奕洋: 实施实验;
李国垒: 数据收集, 论文起草及最终版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: li.yazi@imicams.ac.cn。

[1] 李亚子. sbdrug.xlsx.《国家基本医疗保险、工伤保险和生育保险药品目录(2009 版)》.
[2] 李亚子. gjdrug.xlsx.《国家基本药物目录(2012 版)》.
[3] 李亚子. ah34drug.xlsx. 安徽省新农合药品目录.
[4] 李亚子. fj35drug.xlsx. 福建省新农合药品目录.
[5] 李亚子. hn41drug.xlsx. 河南省新农合药品目录.
[6] 李亚子. js32drug.xlsx. 江苏省新农合药品目录.
[7] 李亚子. jl22drug.xlsx. 吉林省新农合药品目录.
[8] 李亚子. hlj23drug.xlsx. 黑龙江新农合药品目录.
[9] 李亚子. hn43drug.xlsx. 湖南省新农合药品目录.
[10] 李亚子. nmg15drug.xlsx. 内蒙古新农合药品目录.
[11] 李亚子. allWdrug. 国家新农合一体化药品目录(西药).
[12] 李亚子. allCdrug. 国家新农合一体化药品目录(中药).

收稿日期: 2015-10-21
收修改稿日期: 2016-04-08

Building a National System for the Reimbursable Prescription Drugs

Li Yazi Zheng Jianli Zhou Yiyang Li Guolei
(Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100020, China)

Abstract: [**Objective**] This paper examines the current reimbursable prescription drugs list and creates a national prescription drugs catalog for the new rural cooperative medical system (NCMS). [**Methods**] We modified the technology framework of the Unified Medical Language System and used the mapping algorithms to aggregate the multi-source list of reimbursable prescription drugs. [**Results**] We designed the data structure and directory-mapping algorithm, for the integrated NCMS drugs catalog. [**Limitations**] More research was needed to analyze interactions among these drugs. [**Conclusions**] The proposed method helps us develop a list of reimbursable drugs from multiple sources. This new system solves the existing problems of data dictionary aggregation.

Keywords: New rural cooperative medical system Multiple sources dictionary Drug coding Integrated mapping